WHAT HAPPENED TO THE DISTRIBUTION OF INCOME IN SOUTH AFRICA BETWEEN 1995 AND 2001?

Charles Simkins
University of the Witwatersrand
22 November 2004

He read each wound, each weakness clear;
And struck his finger on the place,
And said: 'Thou ailest here, and here".

- Matthew Arnold

1   Introduction

For a country as concerned about income distribution as South Africa is, there is remarkably little national statistical information on household incomes. There are the national accounts, the 1995 and 2000 Income and Expenditure Surveys and the 1996 and 2001 Population Censuses. These sources have limitations. The national accounts tell us nothing about the size distribution of household incomes. The 2000 Income and Expenditure Survey is flawed. The censuses indicate personal incomes in income classes and in both censuses there are well over two million households reporting no income at all. The purpose of this study is to bring the available information within a common framework, proposing corrections to incomplete or erroneous data, in order to make the best judgement possible about changes in income distribution between 1995 and 2001.

2   The national accounts

Table 1 sets out national accounts information from 1994 to 2001 and includes the consumer price index and the population for these years.

The national accounts concept of personal income includes two items not asked for in IES and census questionnaires. These are employer contributions to funds (notably pension and medical aid) and imputed rent. Employer contributions to funds can run as high as 25% for salaried professionals and managers, but are, on average, lower for less skilled formal sector workers and are nothing at all in the case of the self-employed or workers in the informal sector. On average, 10% of compensation of employees is assumed to consist of employer contributions to funds. Included in household expenditure on rent is expenditure on imputed rent (i.e. the implicit return on residential equity owned by occupiers.) 25% of rent paid is assumed to be imputed rent. Subtraction of these two items from the national accounts estimate of current

1

income of households yields 'current comparable income', the comparability being with the IES and censuses. Between 1995 and 2000, current comparable income of households rose by 62.5%. The average real annual rate of growth was 3.3%, while the annual population growth rate was 2.0%, implying a growth of real comparable household income per capita of 1.3%.

3   The income and expenditure surveys

Both income and expenditure surveys had to be re-weighted to recent Statistics South Africa estimates of the population at the time of the surveys. The adjustment was carried out by population group and province. Estimates of population group and provincial totals were obtained using the weights of the original surveys. Population group and provincial multipliers were then applied to the original weights in order to produce the desired new population group and provincial totals, using a method known variously as raking, or the RAS method, or bi-proportional matrices. The age and gender distributions from the surveys were taken simply as they emerged from this process; no further attempt was made to adjust them, even though they deviated somewhat from the age and gender distribution in Statistics South Africa's cohort-component model.

3.1   The 1995 IES

Aggregate household income by population group using the new weights was estimated to be:

	R billion	Share
Africans	127.09	42.0%
Coloureds	22.35	7.4%
Asians	15.40	5.1%
Whites	137.53	45.5%
TOTAL	302.37	100.0%

Of this total R7.30 billion is for alimony and inter-household transfers which cancel out in the national accounts, leaving R 295.07 billion. The IES estimates are for the year ended October 1995 and so should be compared with R 355.92 billion from the national accounts (comparable household income), implying an 82.9% coverage rate.

Three Gini coefficients are calculated from the data. The difference lies between the recipient units. The first coefficient is the most widely used and refers to households. The second coefficient refers to adult equivalents on the OECD scale, counting 1.0 for the first adult, 0.7 each for the second and subsequent adults and 0.5 for each child under 15. The third coefficient also refers to adult equivalents on what is variously known as the modified OECD or EU scale counting 1.0 for the first adult, 0.5 for the second and subsequent adults and 0.3 for children under 15.

	Gini
Households	0.608
OECD adult equivalent	0.642
EU adult equivalent	0.633

3.2   The 2000 IES

Within a short time of its release, it has been clear that a considerable number of observations in the 2000 IES are seriously inaccurate and/or incomplete. It is therefore necessary to remove the worst observations from the data and to reweight the remaining observations.

Ingrid Woolard has categorized observations which are both in the 2000 IES and the Labour Force Survey (LFS 4) to which they were matched. She has regressed the natural logarithm of per capita expenditure and per capita income on the following dependent variables:

- Population group
- Province
- Urban/rural dummy
- The number of adult wage earners divided by the number of adults
- Average adult income from wages
- The number of old age pensioners
- The number of disability grants in the household
- The gender of the household head
- Average educational attainment

The adjusted R^2 for the expenditure equation was 0.548 and for the income equation 0.546. All variables except the disability grant were significant at 5%.

From these regressions, predicted values of the logarithms of per capita expenditure and per capita income were obtained. The standard deviations of the residuals were found to be 0.86 and 0.90 respectively. All records where the reported values were more than two standard deviations from their expected values were identified as outliers.

A sets of flags were created for potentially dirty data. These were assigned as follows:

flag=1 if household income *and* expenditure are outliers and income and expenditure differ by more than one-third;
flag=2 if household expenditure is an outlier and income and expenditure differ by more than one-third;
flag=3 if household expenditure is an outlier and income and expenditure differ by more than one-third;
flag=4 if neither household income nor household expenditure is an outlier but income and expenditure differ by more than one-third;

flag=5 if household expenditure is an outlier but household income and expenditure are within one-third of each other;
flag=6 if household income is an outlier but household income and expenditure are within one-third of each other;
flag=7 if household income *and* expenditure are outliers but household income and expenditure are within one-third of each other.

The following table sets out the distribution of flag values across households.

```
                Flag         |     Freq.     Percent        Cum.
-----------------------------+-----------------------------------
                          OK |    21,101       80.52       80.52
        X&Y outliers&mismatch |       109        0.42       80.94
        X is outliers&mismatch |       229        0.87       81.81
        Y is outliers&mismatch |       336        1.28       83.09
      not outliers but mismatch |     3,954       15.09       98.18
     X is outlier but X&Y match |        97        0.37       98.55
     Y is outlier but X&Y match |        38        0.15       98.69
 X&Y both outliers but match |       342        1.31      100.00
-----------------------------+-----------------------------------
                       Total |    26,206      100.00
```

Households with values of the flag between 1 and 3 (i.e. one or both of the values of income and expenditure an outlier and a mismatch) were dropped from the analysis.

Table 2 sets out a comparison of aggregate incomes by income type between the 1995 and the 2000 IESs. The first point to note is that aggregate measured comparable income rose from R 295.07 in 1995 to R 396.01 billion in 2000, an increase of 34.2%. compared with 62.5% given by the national accounts. Property incomes were reporting as actually falling between 1995 and 2000 and it is apparent that they were very poorly measured in the latter year. Accordingly, upwards adjustments were made to the two main property items: profit from own business and private pensions. The aim was to increase the aggregate income from these two sources to 62.5% above their 1996 levels. This was done using a three-stage procedure:

- Logit regressions were run on the 1995 IES to determine the expected probability that individuals received profits or private pensions. In the case of profits, the independent variables were population group, gender, five year age groups (from 15-19 to 60-64), and whether earnings were received, for the three minority groups (Africans needed no adjustments at this level in 2000). In the case of pensions, the independent variables were population group, gender, age group (from 50-54 up) and whether a state old age pension was received.
- These regressions were then used to predict probabilities of receipt in the 2000 sample on the basis of the relevant characteristics. The expected number of receipts was compared with the actual number of receipts. This created the basis for an adjustment in the probability of receipt among people who reported no receipt and who lived in households where reported income was less than reported expenditure. Allocation of random numbers completed the process of assigning profit or private pension receipts to eligible people who had recorded no such receipt.

- The amount of assigned receipts was determined by regression predictions of amounts among people who reported receipt of profits or private pensions in the 2000 IES. The amount assigned was the predicted value plus a random value selected from a z-distribution multiplied by the standard error of forecast.

Once these procedures had been completed, all property incomes were inflated by a factor sufficient to yield the desired final estimate by population group. For Coloureds, Asians and Whites this was the population group share in 1995, multiplied by the share of the group population in total population in 2000 and divided by the share of the group population in total population in 1995.

This yields an aggregate household income of R 456.65 (excluding alimony and family allowances). A further upward adjustment of incomes by 5.95% brings the total up to the R483.82 billion necessary to reproduce the 1995 coverage rate.

On this basis, the estimate of aggregate income (including alimony and family allowances) by population group in 2000 becomes:

	R billion	Shares
Africans	200.19	40.0%
Coloureds	43.33	8.7%
Asians	26.52	5.3%
Whites	229.87	46.0%
TOTAL	499.91	100.0%

Coloureds, Asians and Whites gained income share between 1995 and 2000 and Africans lost share. The explanation for the adverse trend for Africans must be sought in deteriorating conditions in the lower end of the labour market.

The Gini coefficients, which are:

	Gini
Households	0.669
OECD adult equivalent	0.681
EU adult equivalent	0.677.

These are all considerably higher than the 1995 coefficients, indicating a sizeable increase in inequality between 1995 and 2000. Ignoring current transfers from businesses and transfers from the rest of the world, the proportion of household comparable income deriving from property rose from 27.3% in 1995 to 30.4% in 2000. The proportions of compensation of employees and state transfers dropped accordingly.

The Lorenz curves for the distribution of incomes in 1995 and 2000 appear as Figures 1 and 2.

3.3   A comparison of poverty measures between 1995 and 2000

The starting point for the poverty line is a household income of R 800 per month or R 9 600 per year. This is adjusted by the CPI and household size (the average household size dropped considerably between 1995 and 2000) to give a poverty line of R 7 898 in 1995. Average adult equivalents per household are calculated for the OECD and EU adult equivalent scales and the poverty lines per adult equivalent are calculated in Table 3.

Four poverty measures are calculated in each case. They are:

- The headcount ratio (Foster-Greer-Thorbecke measure 0)
- The poverty gap ratio (Foster-Greer-Thorbecke measure 1)
- The Foster-Greer-Thorbecke measure 2
- The Sen poverty index.

The results are set out in Table 4.

Again the results are unambiguous. On all the poverty measures, poverty increased between 1995 and 2000 against a constant real poverty line per person.

Provincial poverty headcount ratios are calculated for household income and are reported in Table 5. They are generally what one would expect with the Eastern Cape and Limpopo reporting the highest ratios and Gauteng and Western Cape reporting the lowest values. The standard errors on some of the estimates are quite high. In five of the nine provinces, we can be reasonably sure that the headcount ratio rose between 1995 and 2000. In three others, the probable movement is downwards. In one province no clear movement can be discerned.

4   The population censuses

Income information was not complete in either the 1996 or the 2001 censuses. In the 1996 census there was no imputation of income. Of 9 062 348 households, 1 141 665 (12.6%) were returned as having no income and another 1 070 289 (11.8%) were returned as having missing incomes for one or more members. There was imputation of income in the 2001 census. Nonetheless, 2 732 107 (23.2%) of 11 770 270 households were returned as having no income. Clearly imputation/further imputation of income is needed in both censuses. This imputation was carried out on the 10% samples of each census.

Since income data in the population censuses are grouped into categories, the Gini coefficient and the poverty coefficients were calculated using the World Bank's POVCAL program for grouped data.

**4.1	The 1996 population census**

For the calculation of aggregate household income, mean income per income class was always taken to be:

- Two-thirds of the upper limit of the first non-zero income category
- The mean of the lower and upper limits of all other categories except the top (open) category
- Double the lower limit of the top category.

The estimate of aggregate household income as presented in the 10% sample was R 272.8 billion, well short of the R302.4 billion measured a year before in the 1995 IES. The imputation proceeded in the following stages:

- Imputation of old age pension income, when people of pensionable age (men 65 and over, women 60 and over) live in households with zero reported income. The imputed incomes were distributed according to the reported incomes of people of pensionable age in the same population group, with the same gender and education.
- Imputation of incomes for the employed in households with no income for men between 15 and 64 and women between 15 and 59. The imputed incomes were distributed according to the reported incomes of people in the population group, with the same gender and in the same age group and educational group.
- Imputation of incomes for the ill/disabled in households with no income.
- Imputation of incomes for the unemployed and the people who could not find work in households with no income.
- Imputation of incomes for scholars and students, home-makers, pensioners and those choosing not to work in households with no income.
- Imputation of incomes to the last remaining people in households with no income.

These procedures may often have assigned more than one income to households which had no income before imputation, so that all but the lowest imputed income in each such household was culled from the data set. This puts imputation on the most conservative possible footing.

After the imputation, aggregate household income becomes:

	R billion	Share
African	147.02	43.6%
Coloured	26.29	7.8%
Asian	14.63	4.3%
White	149.54	44.3%
TOTAL	337.48	

This aggregate is 11.6% up on the 1995 IES estimate a year earlier, and compares well with the 12.5% increase in comparable current income from the national

accounts. The shares by population group are close as well. However, there are some important differences:

- The mean household size in the 1996 10% sample was 3.76 compared with 4.40 in the 1995 IES.
- The Gini coefficient in the 1996 10% sample as amended was 0.660, compared with 0.608 in the 1995 IES.
- The poverty line was set at R 7 240 per annum for a household in the 1996 census. This is derived from the 2000 poverty line, with adjustments for the CPI and average household size. On this basis, the poverty measures become:

  Head count   0.346
  Poverty gap   0.168
  FGT2   0.102

  These are all considerably higher than the measures derived from the 1995 IES.

The 1995 IES sample was drawn from people living in households. The 1996 Census sample included people living in collective living quarters and institutions as well as people living in households. A second run using just the people living in households in the 1996 census yielded an average household size of 4.15 with a Gini coefficient of 0.661 and higher poverty measures (head count: 0.661, poverty gap 0.192 and FGT2 0.122).

Table 6 compares the distribution of individual incomes between the 1995 IES and the 1996 Census. The 1996 estimates are found by linear interpolation within income classes. Table 6 shows that the census (households only) as adjusted found 41.0% of the population to be in receipt of an income, compared with only 35.4% in the Income and Expenditure Survey. On the other hand, the real census incomes at the various percentiles were well below the IES incomes, especially at the 10th and 20th percentiles. The difference was smallest at the 90th, 95th and 99th percentiles. This explains the higher Gini coefficient and higher poverty measures in the census.

4.2 The 2001 Population Census

The imputation procedures used on the 1996 census were applied to the 2001 census as well. The resulting estimate of household income was R625.85 billion, considerably higher than the estimate from the 2000 IES. The main reason is an overly heavy upper tail of the distribution of individual incomes in the 2001 census.

Table 7 compares incomes at various percentiles between the 2000 IES and the 2001 census. Again, incomes at the first three deciles in the Census distribution are well below those in the IES distribution. But the main point to note is the increasing exaggeration of incomes in the census distribution above the 93rd percentile. There was a tendency to check boxes at too high a level in the top six income classes in the census.

Accordingly, individual incomes in the top six classes were re-arranged to reproduce the upper tail of the 2000 IES, with an adjustment of incomes for the time difference. This was achieved by moving an appropriate proportion of incomes down an income class, starting from the top. After this adjustment, aggregate household income fell to R 576.79 billion, distributed across the population groups as follows:

	R billion	Share
Africans	243.09	43.8%
Coloureds	47.84	8.4%
Asians	27.31	4.9%
Whites	236.83	42.7%
TOTAL	555.07	100.0%

The shares by population group are higher for Africans and lower for everyone else when the 2001 census is compared with the 2000 IES.

The Gini coefficient for the distribution of household incomes in the 2001 Census is 0.692, higher than the 0.669 found by the 2000 IES. The poverty line after adjustment for the CPI and household size is R 10 189 per annum and the poverty measures are:

Headcount	0.365
Poverty gap	0.156
FGT2	0.083

The headcount measure is higher than in the 2000 IES, the poverty gap measure is the same, and the FGT2 measure is lower.

5   Conclusion

The Gini coefficient rose from 0.608 to 0.669 between the 1995 and the 2000 IES, as adjusted. It rose from 0.660 to 0.692 between the 1996 and the 2001 censuses, as adjusted. At the bottom of the distribution, there is evidence of better correspondence between the 2001 census and the 2000 IES than between the 1996 census and the 1995 IES. Incomes at the bottom end of the distribution were more understated in 1996 than in 2001. This introduces an upward bias in the Gini coefficient obtained from the 1996 census. The conclusion is clear: All the evidence we have, suitably interpreted, indicates that inequality, as measured by the Gini coefficient, increased by a substantial margin between 1995 and 2001.

The evidence on poverty based on a fixed real poverty line per capita is more mixed (R 9 600 per year in 2000 for the average household size of 3.87 persons), since real per capita household income increased over the period. Comparison of four poverty measures between the 1995 and 2000 IES show a clear increase. Comparison of poverty measures between the 1996 and 2001 censuses show a mixed picture. However, in the light of the finding the low incomes were more understated in the 1996 census than in the 2001 census, the 1996 census poverty measures are upwardly

biased, and so the increase between 1996 and 2001 is downwardly biased. On balance, it is likely that poverty has worsened as well.

Tables:

Table 1 - Current income of households
R millions

	1994	1995	1996	1997	1998	1999	2000	2001
Compensation of employees	240416	272916	306225	338152	370454	395421	421168	453943
Income from property	94052	110433	121666	142160	148752	165799	193703	223429
Current transfers from government	17357	18452	24203	27145	29032	32285	35870	38311
Current transfers from businesses	305	400	449	503	588	634	674	705
Transfers from the rest of the world	83	110	105	155	190	231	260	292
Current income	**352213**	**402311**	**452648**	**508115**	**549016**	**594370**	**651675**	**716680**
Less employers contribution to funds	24042	27292	30623	33815	37045	39542	42117	45394
Less imputed rent	6693	7624	8624	9645	10668	11624	12491	13694
Current comparable income	**321478**	**367396**	**413402**	**464655**	**501303**	**543204**	**597067**	**657592**
Consumer price index	66.6	72.4	77.7	84.4	90.2	94.9	100.0	105.7
Current comparable income (2000 prices)	482700	507453	532048	550539	555768	572396	597067	622130
Growth		5.1%	4.8%	3.5%	0.9%	3.0%	4.3%	4.2%
Population (million)		39.7					43.8	

Sources: Reserve Bank Quarterly Bulletin June 2002
South African Statistics 2002
Statistics South Africa demographic model

Table 2 - Income components, IES 1995 and IES 2000
R billion

	1995 Africans	Coloureds	Asians	Whites	1995	2000 Africans	Coloureds	Asians	Whites	2000	Per cent increase	Revised total	
Salary	91.342	17.775	9.790	85.387	204.294	135.008	31.195	15.101	113.476	312.768	53.1%		
Bonus	1.860	0.401	0.257	1.724	4.242	3.839	0.835	0.618	4.461	10.786	154.3%		
Commission	0.093	0.041	0.048	0.902	1.084	0.267	0.029	0.094	1.008	1.792	65.3%		
Part time	0.861	0.084	0.055	0.593	1.593	1.120	0.140	0.075	0.982	2.393	50.2%		
Profit	12.825	1.470	4.262	32.554	51.111	10.629	0.687	0.813	10.474	23.091	*-54.8%*	83.055	
Property	0.120	0.006	0.100	0.505	0.731	0.177	0.020	0.000	1.416	1.697	132.1%	1.697	
Other						0.382	0.062	0.091	0.446	1.106		1.106	
Royalty	0.032	0.000	0.004	0.038	0.074	0.165	0.005	0.000	0.059	0.280	278.4%	0.280	
Interest	0.358	0.050	0.086	2.263	2.757	0.476	0.047	0.051	1.261	2.111	-23.4%	2.111	
Dividends	0.101	0.005	0.034	0.857	0.997	0.135	0.070	0.154	0.727	1.096	9.9%	1.096	
Pension	2.250	0.563	0.276	9.573	12.662	1.998	0.824	0.283	7.543	12.421	*-1.9%*	20.576	
Annuity	0.144	0.021	0.004	0.755	0.924	0.316	0.048	0.074	1.907	2.949	219.2%	2.949	
State old age pension	9.108	0.927	0.195	1.189	11.419	11.894	1.182	0.553	2.338	16.946	48.4%		
Disability grant	1.110	0.418	0.070	0.188	1.786	2.001	0.566	0.145	0.373	3.170	77.5%		
Family grant	0.630	0.213	0.058	0.091	0.992	0.845	0.231	0.066	0.309	1.583	59.6%		
Workmen's Compensation	0.260	0.025	0.021	0.095	0.401	0.974	0.108	0.049	0.632	1.820	353.9%		
Alimony	0.927	0.149	0.030	0.382	1.488	2.004	0.253	0.117	0.490	3.057	105.4%		
Allowance	5.068	0.204	0.114	0.439	5.825	9.526	0.493	0.232	1.398	12.126	108.2%		
Total	127.089	22.352	15.404	137.535	302.380	181.756	36.795	18.516	149.300	411.192	36.0%		
Earnings	94.156	18.301	10.150	88.606	211.213	140.234	32.199	15.888	119.927	327.739	55.2%	327.739	55.2%
Property	15.830	2.115	4.766	46.545	69.256	14.278	1.763	1.466	23.833	44.751	-35.4%	112.870	63.0%
Transfers	11.108	1.583	0.344	1.563	14.598	15.714	2.087	0.813	3.652	23.519	61.1%	23.519	61.1%
Sub-total	121.094	21.999	15.260	136.714	295.067	170.226	36.049	18.167	147.412	396.009	34.2%	464.128	57.3%

Table 3 - Poverty lines 1995 and 2000

	1995	2000
Poverty line for average household size (2000)	7904	9600
CPI	72.4	100.0
Average household size	4.402	3.871
Average OECD adult equivalents	3.096	2.758
Average EU adult equivalents	2.416	2.184
OECD poverty line per adult equivalent	2553	3481
EU poverty line per adult equivalent	3271	4396

Table 4 - Poverty measures from the 1995 and 2000 IESs

	1995	2000
Income per household		
Headcount ratio	0.293	0.338
Poverty gap ratio	0.134	0.156
FGT 2 measure	0.080	0.099
Sen measure	0.176	0.210
Income per OECD adult equivalent		
Headcount ratio	0.337	0.380
Poverty gap ratio	0.155	0.181
FGT 2 measure	0.095	0.115
Sen measure	0.206	0.241
Income per EU adult equivalent		
Headcount ratio	0.320	0.360
Poverty gap ratio	0.146	0.168
FGT 2 measure	0.089	0.106
Sen measure	0.194	0.225

Table 5 - Headcount measures of poverty by province, 1995 and 2000

	1995	**2000**	**2000** 95% confidence interval		
Eastern Cape	0.452	0.527	Up	0.500	0.553
Free State	0.494	0.409	Down	0.369	0.449
Gauteng	0.109	0.201	Up	0.182	0.219
KwaZulu-Natal	0.247	0.353	Up	0.328	0.377
Limpopo	0.436	0.515	Up	0.492	0.539
Mpumalanga	0.270	0.366	Up	0.332	0.400
Northern Cape	0.402	0.346	Down	0.297	0.395
North West	0.406	0.354	Down	0.324	0.385
Western Cape	0.130	0.125	Same	0.111	0.146

Table 6 - Income sizes and percentages of people receiving income

	IES 95	Census 96	Census/IES*
Percentage of people **in receipt of an income**	35.4	43.0	
Individual income level at:			
(Rand per annum)			
10th percentile	2725	2104	69%
20th percentile	4720	3495	66%
30th percentile	4920	4769	86%
40th percentile	6240	6072	86%
50th percentile	9840	9689	88%
60th percentile	14400	13528	84%
70th percentile	19200	17867	83%
80th percentile	30000	28054	83%
90th percentile	48000	48534	90%
95th percentile	70873	68619	86%
99th percentile	150000	164183	97%

* Includes an adjustment for the difference in date

Table 7 - Individual income levels at various percentiles, IES 2000 and Census 2001

	IES 2000	Census 2001	Census/IES *
Percentage of people **in receipt of an income**	0.387	0.440	
Percentile	**Individual income level (Rand per annum)**		
10	2495	2103	77%
20	4450	4206	86%
30	6357	5882	84%
40	6866	7390	98%
50	8476	8897	95%
60	12714	12491	89%
70	19071	17909	85%
80	29420	31018	96%
90	54670	62896	104%
95	83167	112838	123%
99	204418	331084	147%

* Includes an adjustment for the difference in date

Figures

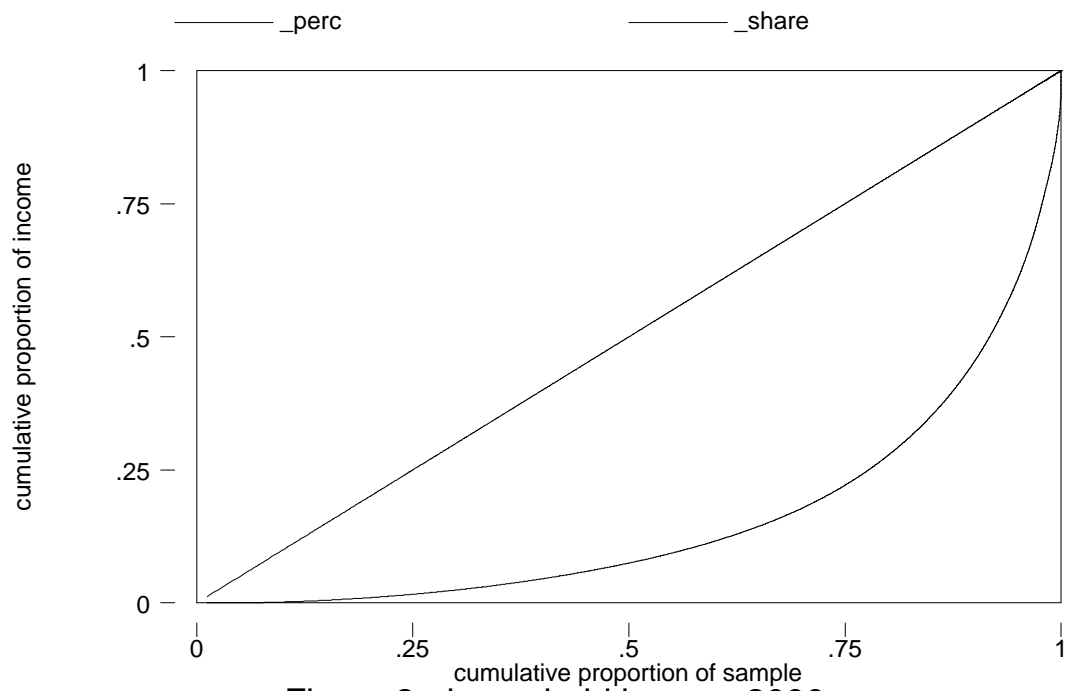_perc   _share
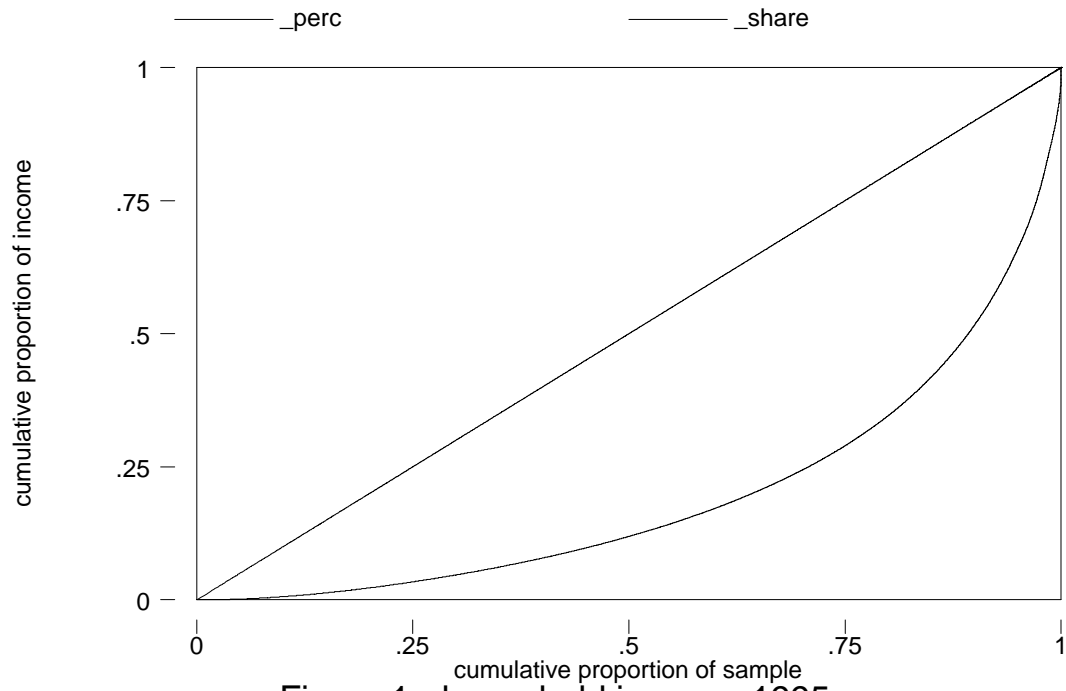
Figure 1 - household income, 1995

_perc   _share

Figure 2 - household income 2000